

Bridging Platforms: Integrating RStudio POSIT and SAS Grid in the same environment

Sebastià Barceló, sebastia.barcelobauza@cytel.com, Cytel Inc, Geneva, Switzerland
Malte Stein, malte.stein@cytel.com, Cytel Inc, Berlin, Germany

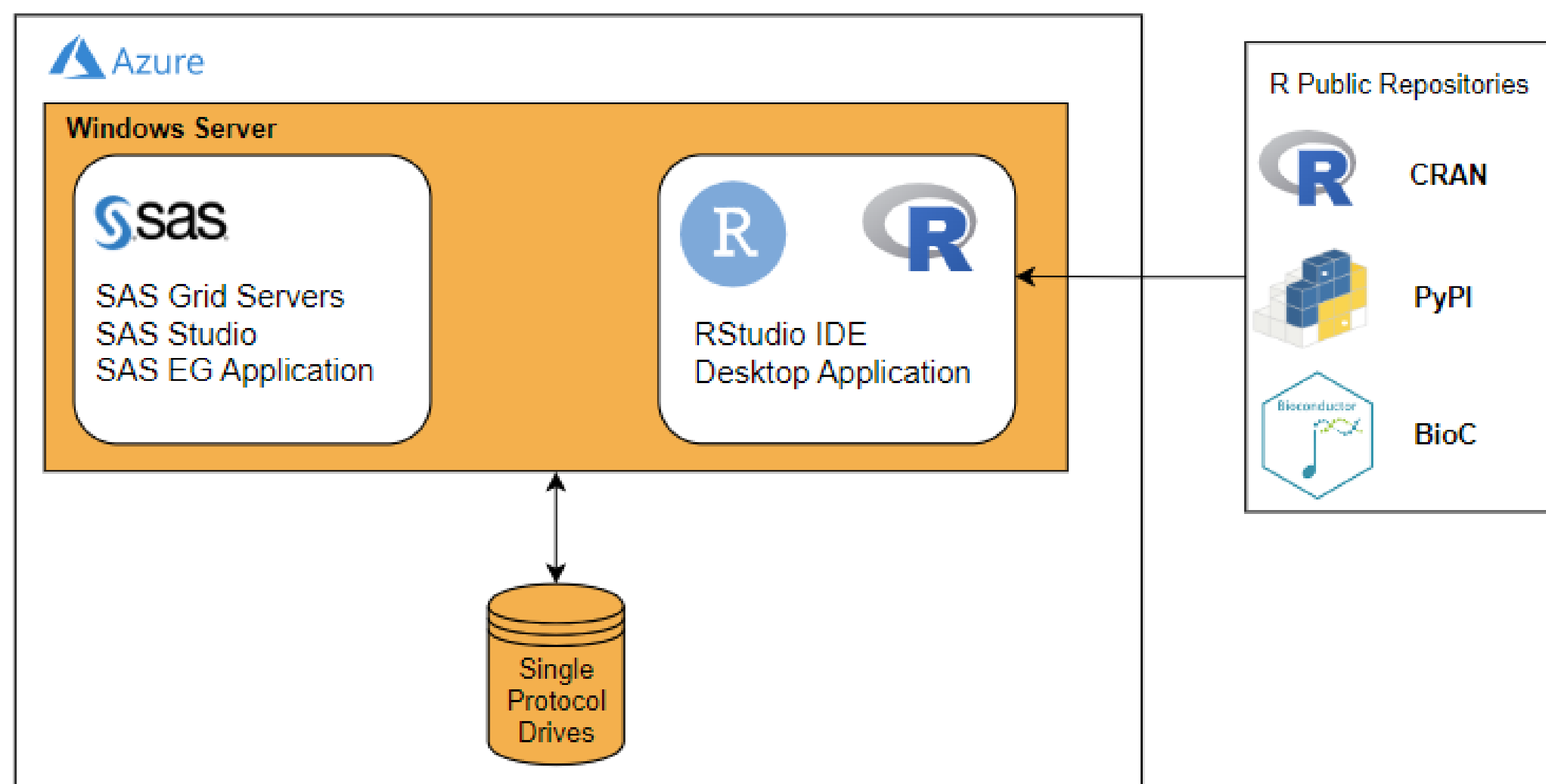
Abstract

Integrating RStudio POSIT into the same environment where we have SAS Grid posed several challenges due to disparate platforms - SAS Grid installed on Windows server and POSIT installed on Linux server.

This presentation outlines the hurdles encountered during the integration process and the innovative solutions devised to overcome them. From addressing compatibility issues to streamlining accessing to the same drives from both platforms, we navigated through complexities to ensure seamless integration.

Through collaborative efforts and strategic planning, we successfully bridged the gap between Windows and Linux environments, enabling efficient utilization of RStudio POSIT and SAS Grid accessing the same data. Join us to explore our journey, learn from our experiences, and gain insights into effectively integrating diverse technologies within a unified environment.

1. Original Infrastructure



2. New POSIT tools to incorporate

Workbench

Analyze, develop & publish

- IDE for data scientists to analyze data, develop tools, and create content for sharing.

Package Manager

Curate, distribute & secure

- Platform to manage, distribute, and secure R and Python packages across teams and environments.

Connect

Deploy, collaborate & scale

- Platform to deploy, collaborate on, and scale Shiny apps, dashboards, APIs, and reports for sharing insights.

3. Different instances of Connect for Internal vs External access

Pros:

- **Security:**
 - Access Control: Easier to apply different security policies and user authentication.
 - Data Segregation: Internal data and content are isolated from external access, minimizing the risk of accidental exposure of sensitive information
- **Performance:** Dedicated resources to reduce the risk of traffic created by external users impacts on internal users and vice versa.
- **Compliance:** Keeping internal and external content separate might help meet compliance obligations, especially when external content must meet different legal or regulatory standards (e.g. use of different packages due to their licenses).

Cons:

- Increased maintenance and higher costs
- Complex user management
- Integration challenges, more complex workflow to publish content in the External Connect (cannot be done directly from the single instance of Workbench)...

4. Data Storage and Workbench access

Related to data storage

- To ensure seamless data access between SAS Grid (Windows) and POSIT (Linux), we implemented **dual protocol drives** supporting both SMB (Windows) and NFS (Linux), resolving compatibility issues and enabling efficient cross-platform data sharing.
- Migrating to dual protocol drives also posed challenges with **folder permissions**, particularly around mapping Active Directory (Windows) and Linux-based file permissions at the subfolder level. Due to this, we kept using single protocol drives during some time, but the plan is to decommission those drives after completing the migration.

Workbench access

- During the integration, we encountered an issue with the **client_init** process, which caused unstable access to POSIT Workbench. To resolve this, we implemented an automatic Chrome cache cleaning process, which stabilized the connections and ensured reliable access to the shared drives.

5. Integration Challenges

RStudio – Package Manager

- RStudio Desktop connects seamlessly to the Package Manager, with no Windows-Linux compatibility issues in this setup. However, we encountered cache-related challenges during package installation for RStudio Desktop, which required additional troubleshooting.

RStudio/POSIT – SAS Grid

- One limitation is the inability to submit SAS jobs directly from POSIT Workbench. Currently, SAS jobs can only be submitted via RStudio Desktop, posing a constraint in the integrated workflow if we want to create Shiny applications creating and executing SAS code.

POSIT – Git

- We've experienced an issue with triplicated sessions from the **Git Dual Protocol drive**, which is still under investigation by POSIT. Resolving this will help streamline the integration with version control.
- To publish applications for external access without having POSIT Workbench in that server, we decided to use the **git-backed** process, which allows you to publish content in connect directly from a git repository. With this process, POSIT Connect will deploy the content automatically when new commits are pushed.

6. Final Structure

